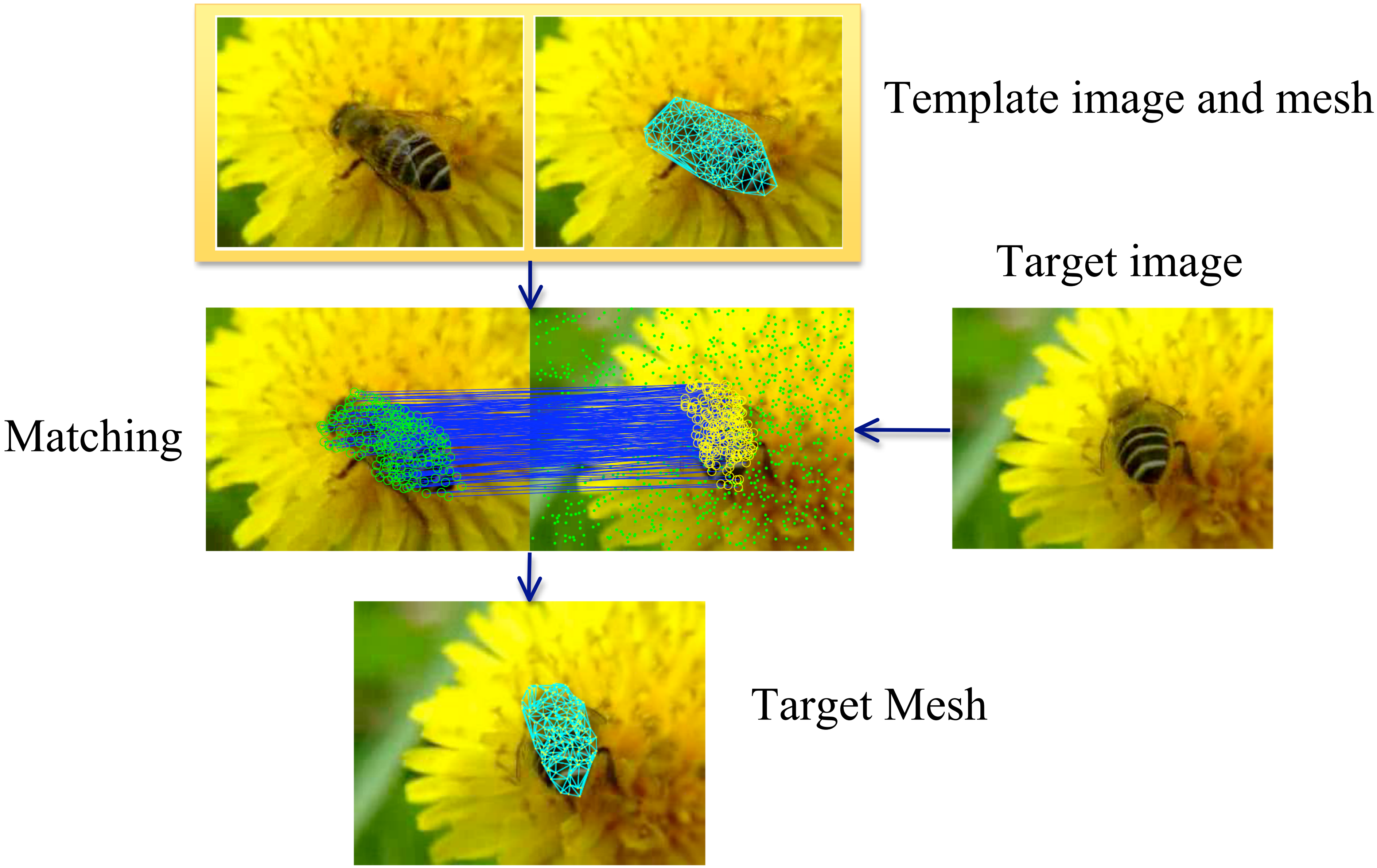
Linear Solution to Scale and Rotation Invariant Object Matching

Hao Jiang and Stella X. Yu

Computer Science Department

Boston College

Problem

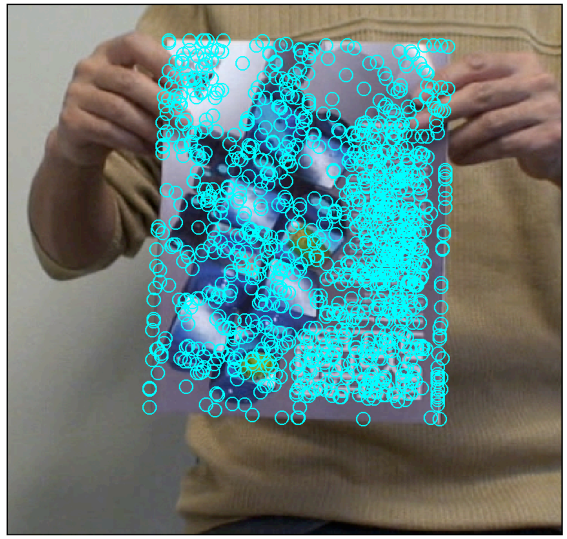


Challenges

Template Image

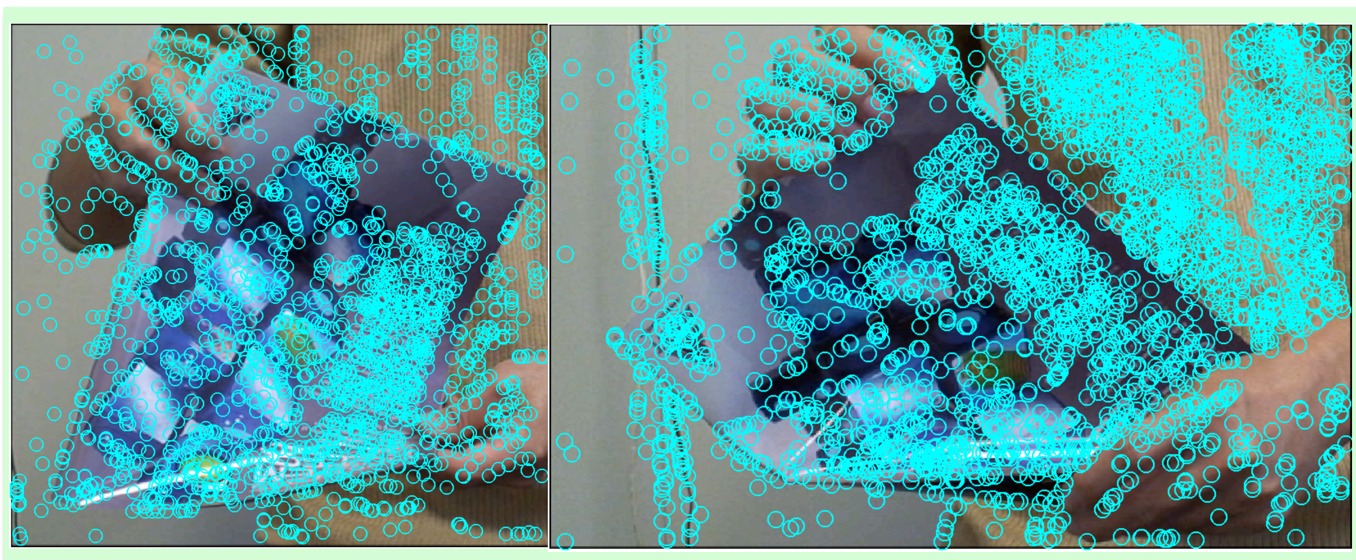
Target Images

Challenges



Template Image

Target Images



Some Related Methods

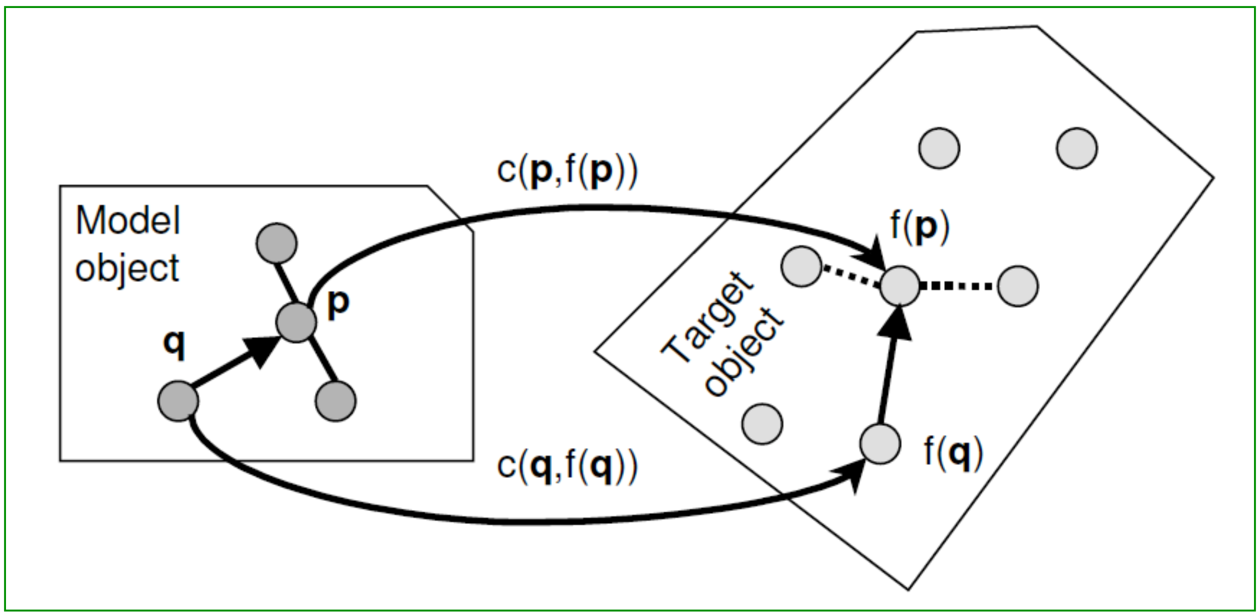
- Hough Transform and RANSAC
- Graph matching
  - Dynamic Programming
  - Max flow – min cut [Ishikawa 2000, Roy 98]
  - Greedy schemes (ICM [Besag 86], Relaxation Labeling [Rosenfeld 76])
  - Back tracking with heuristics [Grimson 1988]
  - Graph Cuts [Boykov & Zabih 2001]
  - Belief Propagation [Pearl 88, Weiss 2001]
- Convex approximation [Berg 2005, Jiang 2007, etc.]

The Outline of the Proposed Method

- A linear method to optimize the matching from template to target
- The linear approximation is simplified so that its size is largely decoupled from the number of target candidate features points
- Successive refinement for accurate matching

The Optimization Problem

$$\min_{\mathbf{f},s,R}\{\underbrace{\sum_{\mathbf{p}\in\mathcal{M}} c(\mathbf{p},\mathbf{f}(\mathbf{p}))}_{\text{Matching cost}}+\lambda \underbrace{\sum_{\{\mathbf{p},\mathbf{q}\}\in\mathcal{N}} \|R\cdot(\mathbf{p}-\mathbf{q})-s\cdot(\mathbf{f}(\mathbf{p})-\mathbf{f}(\mathbf{q}))\|}_{\text{Pairwise consistency}}\}$$



In Compact Matrix Form

$$\min\ (X, s, R) = \underbrace{\mathrm{tr}(C'X)}_{\text{Matching cost}} + \underbrace{\lambda|E(MR - sXT)|}_{\text{Pairwise consistency}}$$

$$\text{subject to}\ \ X1_{n_t} = 1_{n_m}, X \in \{0,1\}^{n_m \times n_t}$$

$$s > 0$$

$$R'R = I$$

We will turn the terms in circles to linear approximations

The L1 Norm Linearization

- It is well known that L1 norm is "linear"

$$\min |x| \Longleftrightarrow \begin{array}{l} \min (y + z) \\ \text{Subject to } x = y - z \\ \qquad\qquad y, z >= 0 \end{array}$$

- By using two auxiliary matrices Y and Z,

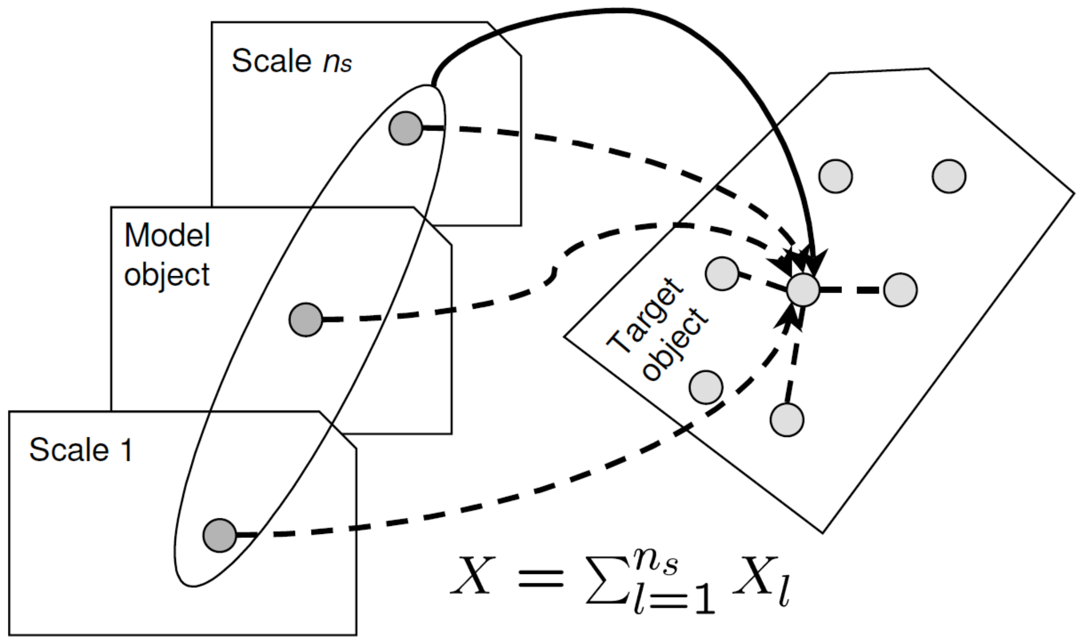
$$|\,EMR - sEXT\,| \rightarrow 1'_{ne}(\,Y + Z\,)\,1_2$$

$$\text{subject to } \mathrm{Y} - \mathrm{Z} = EMR - sEXT$$

$$Y, Z >= 0$$

Consistency term

Linearize the Scale Term



$$X = \sum_{l=1}^{n_s} X_l$$

$$sX = \sum_{l=1}^{n_s} s_l X_l \quad \Rightarrow \quad sEXT = \sum_{l=1}^{n_s} s_l E X_l T$$

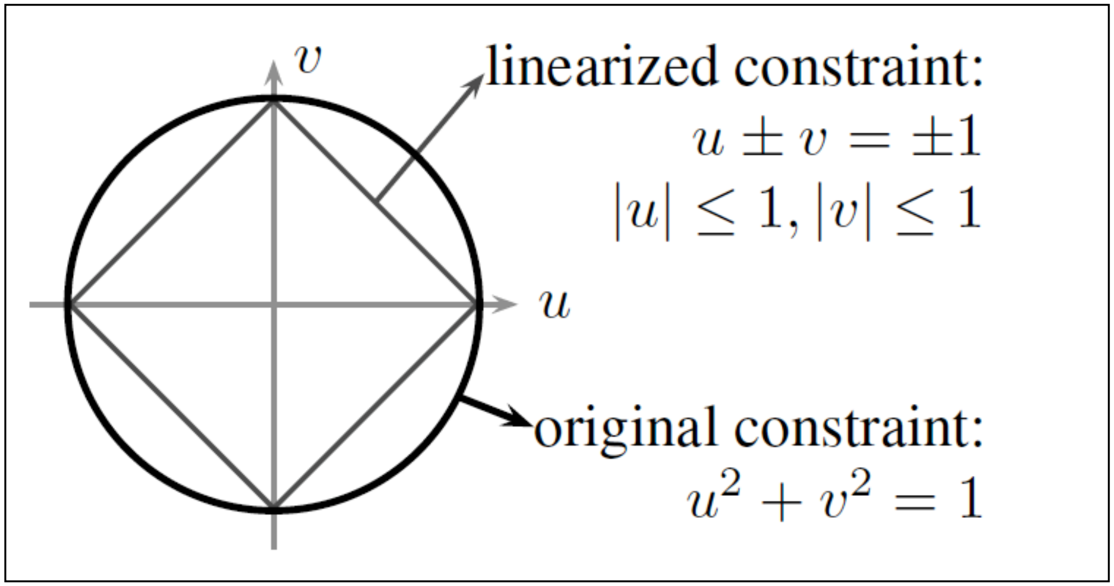
All sites select the same scale $\quad \sum_{l=1}^{n_s} s_l X_l 1_{n_t} = s 1_{n_m}$

Linearize Rotation Matrix

$$R'R = I \;\Rightarrow\; R = \begin{bmatrix} u & -v \\ v & u \end{bmatrix}$$

$$u \pm v = \pm 1, \quad |u| \leq 1,\ |v| \leq 1$$

In graphics:



The Linear Optimization

$$\text{LP:} \quad \min \varepsilon(X, s, u, v, Y, Z, X_1, \ldots, X_{n_s}) = \operatorname{tr}(C'X) + \lambda 1'_{n_e}(Y+Z)1_2$$

$$\text{subject to } Y - Z = EM \begin{bmatrix} u & -v \\ v & u \end{bmatrix} - \sum_{l=1}^{n_s} s_l E X_l T$$

$$Y, Z \geq 0, \quad u \pm v = \pm 1, \quad |u| \leq 1, \quad |v| \leq 1$$

$$X = \sum_{l=1}^{n_s} X_l, \quad X_l \geq 0, \forall l$$

$$\sum_{l=1}^{n_s} s_l X_l 1_{n_t} = s 1_{n_m}$$

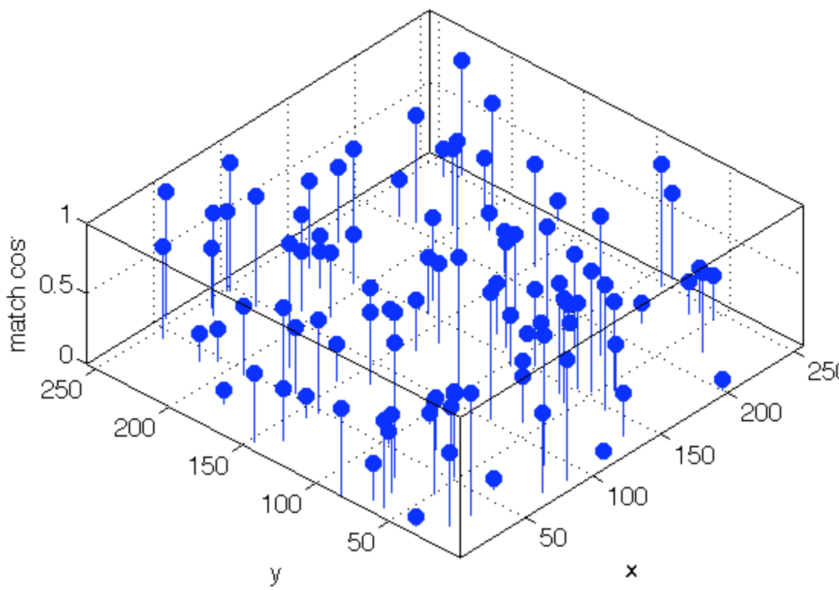
$$X 1_{n_t} = 1_{n_m}, X \geq 0$$

Target point estimation:

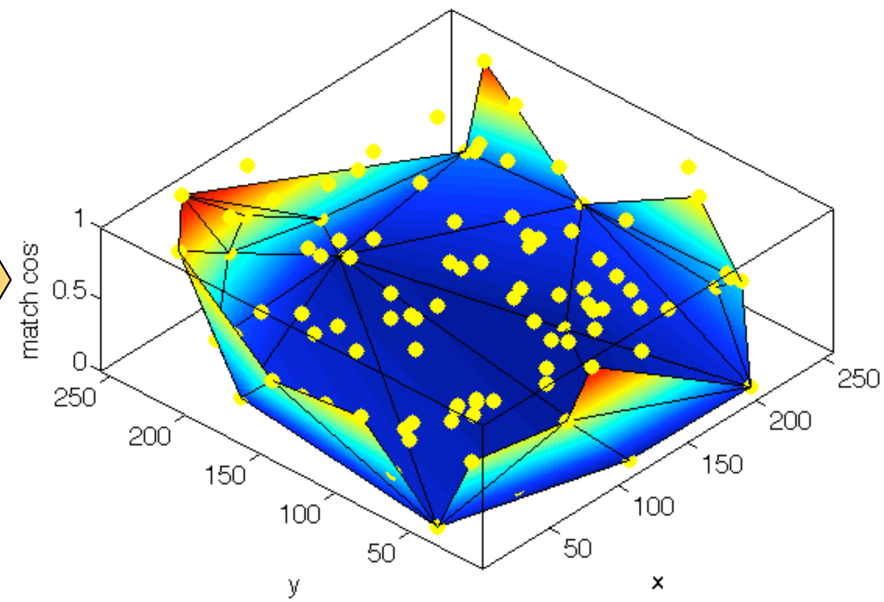
$$T^* = XT$$

The Lower Convex Hull Property

- Cost surfaces can be replaced by their lower convex hulls without changing the LP solution



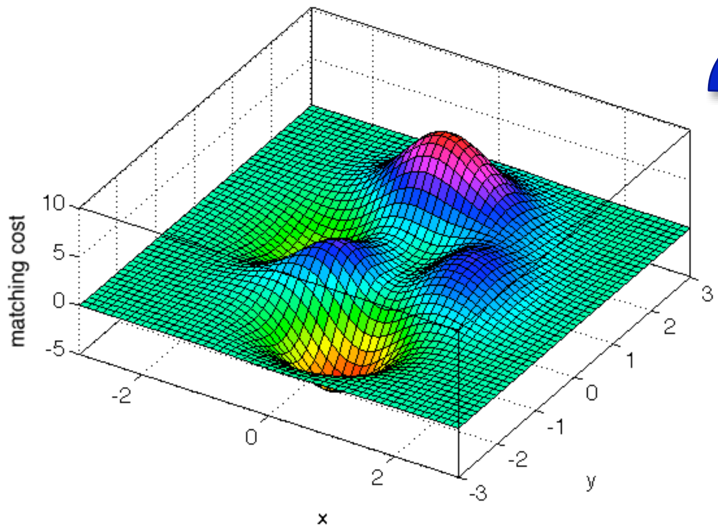
A cost "surface" for one point on the template

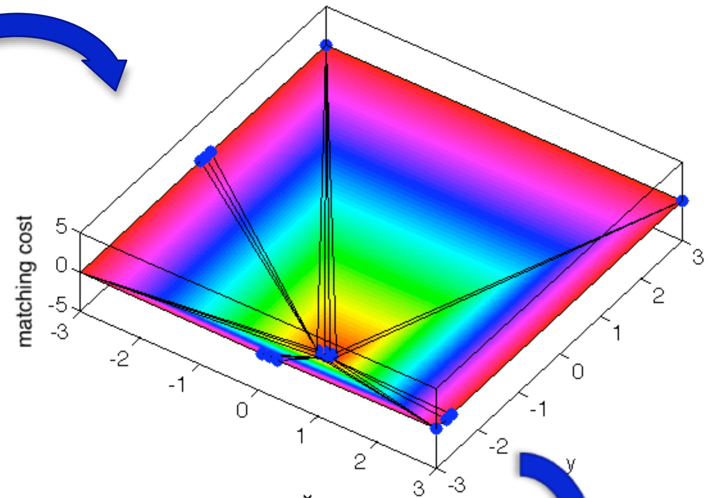


The lower convex hull

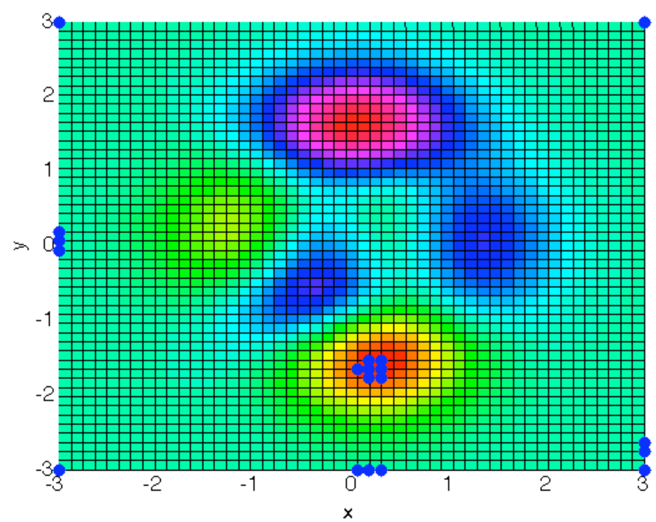
Removing Unnecessary Variables

- Only the variables that correspond to the lower convex hull vertices need to be involved in the optimization





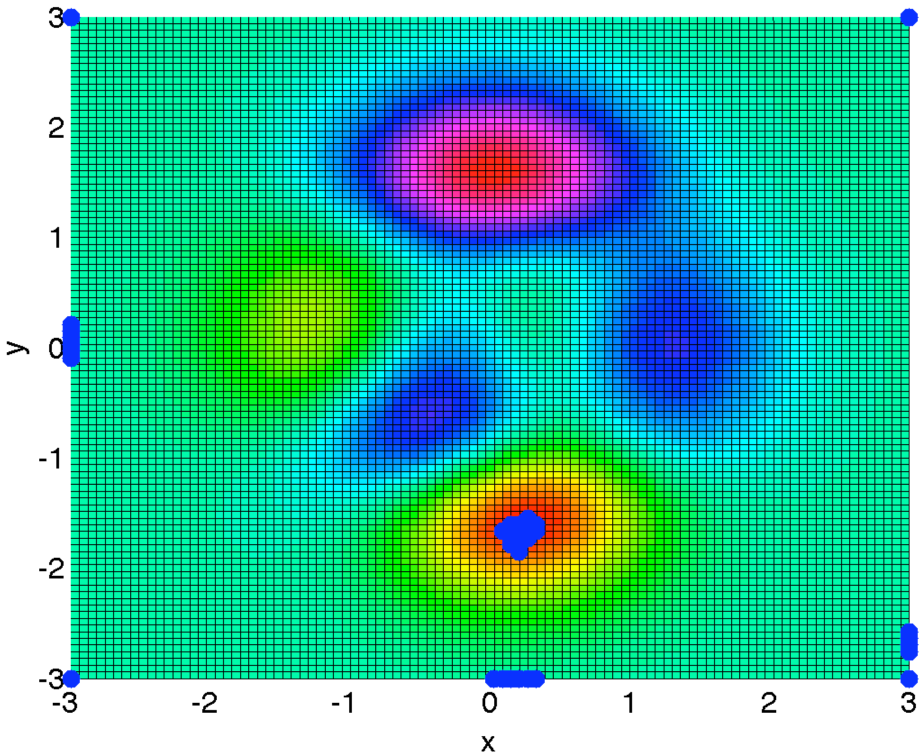
# of original target candidates: 2500



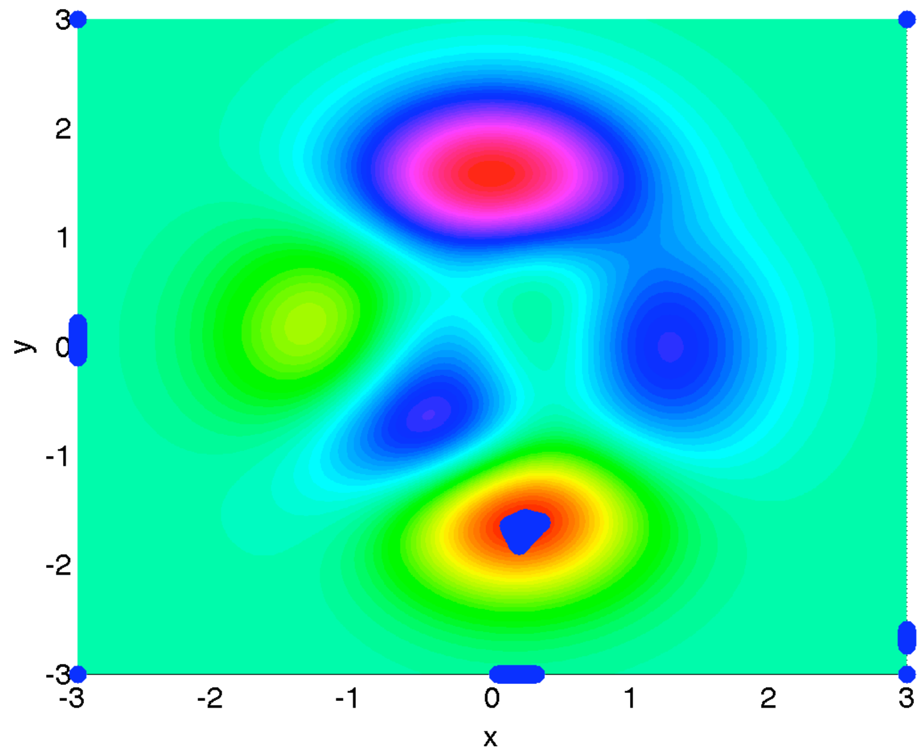
# of effective target candidates: 19

Complexity of the LP

10,000 target candidates



1 million target candidates

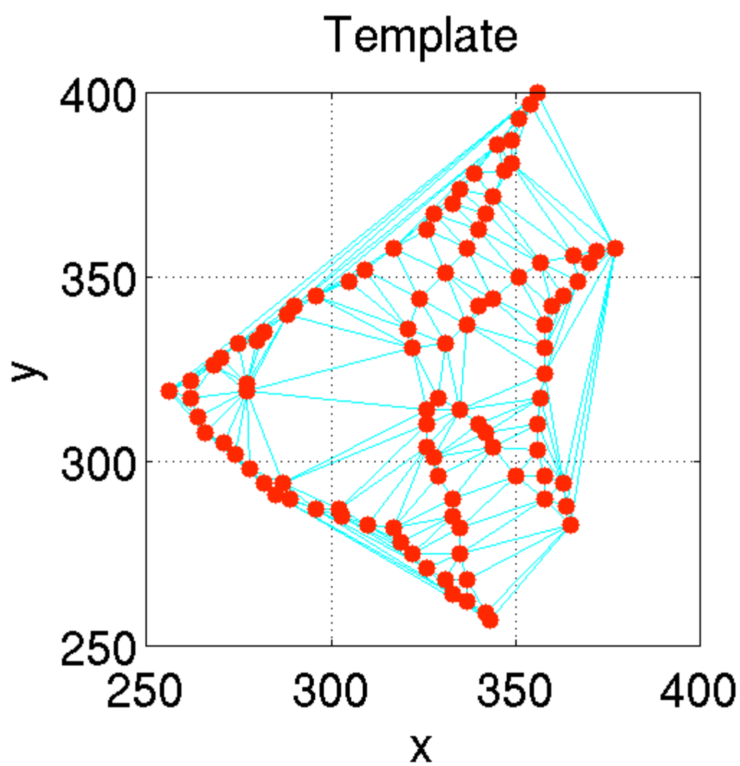


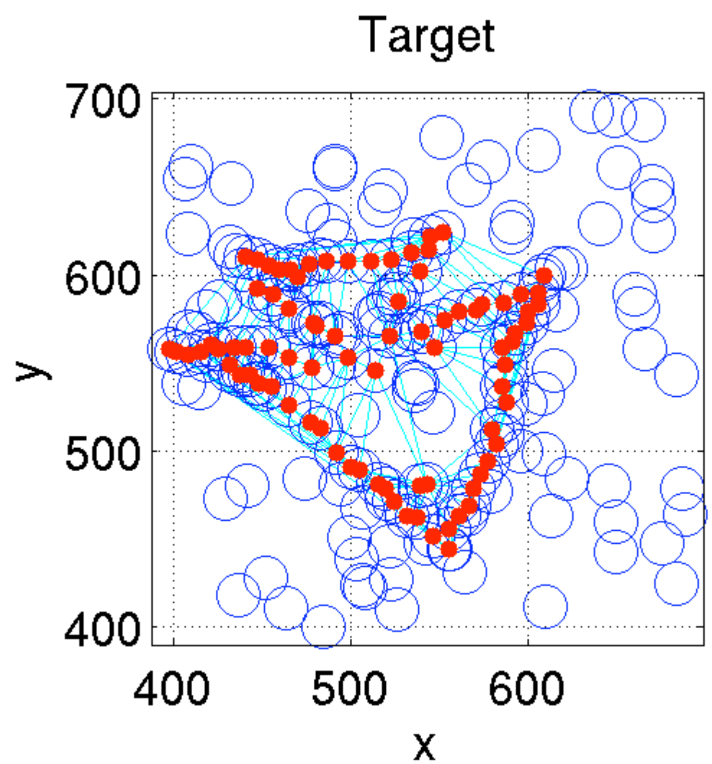
28 effective target candidates

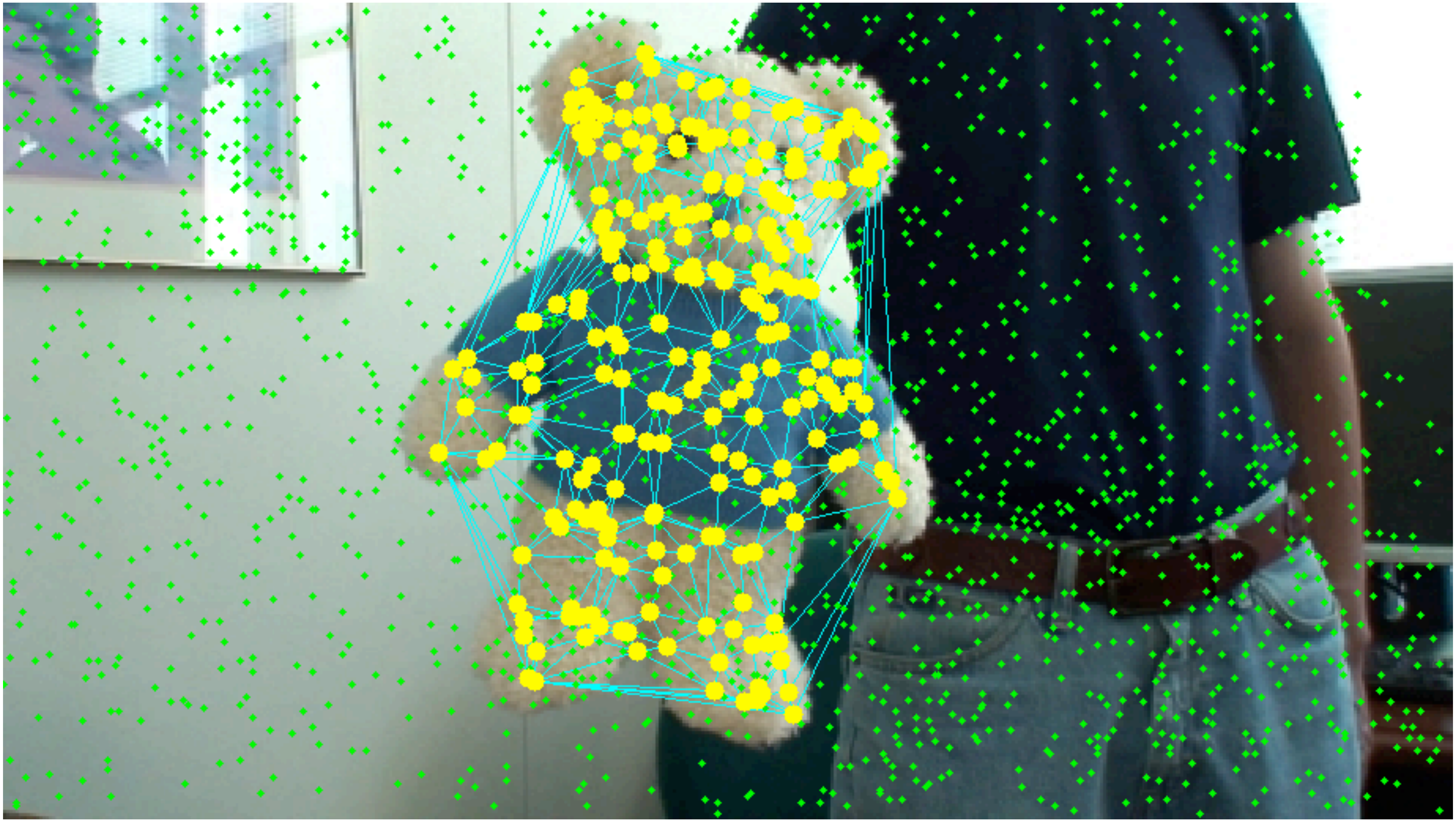
1034 effective target candidates

The size of the LP is largely decoupled from the number of target candidates

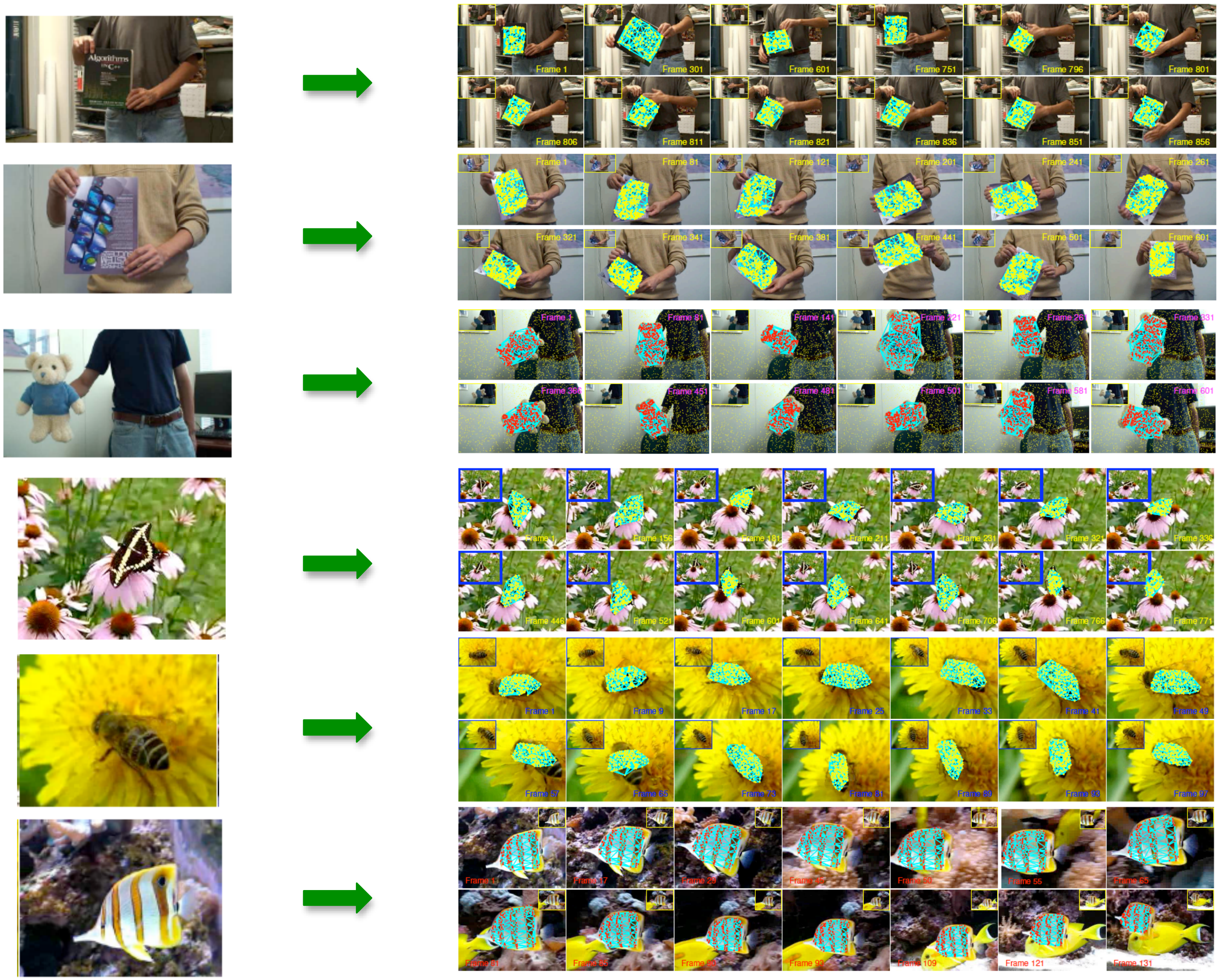
Successive Refinement



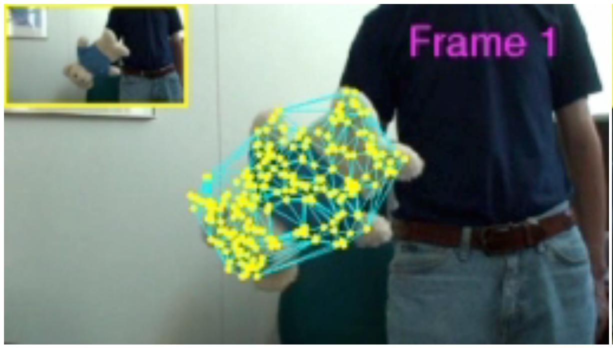


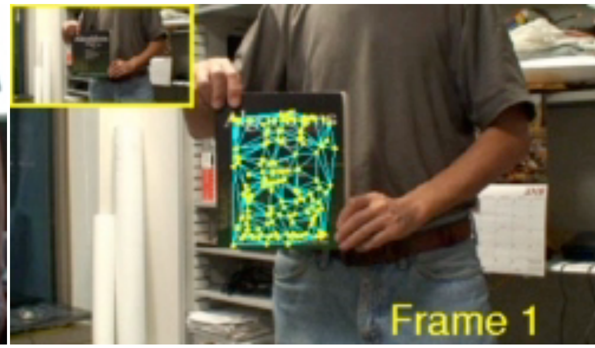


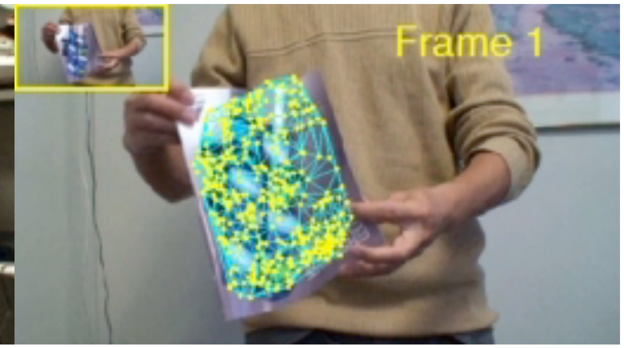
Matching Objects in Real Images

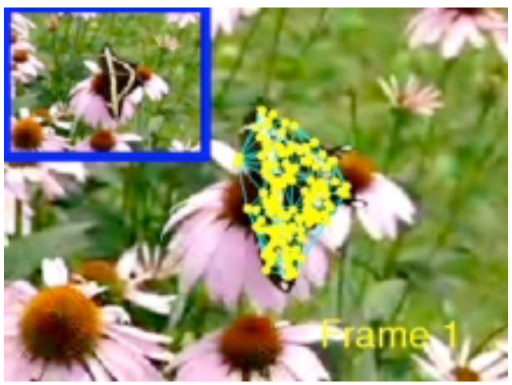


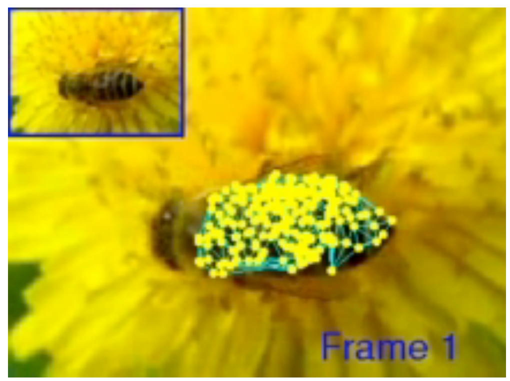
Result Videos

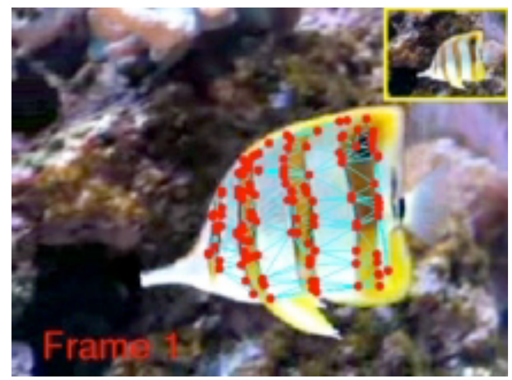








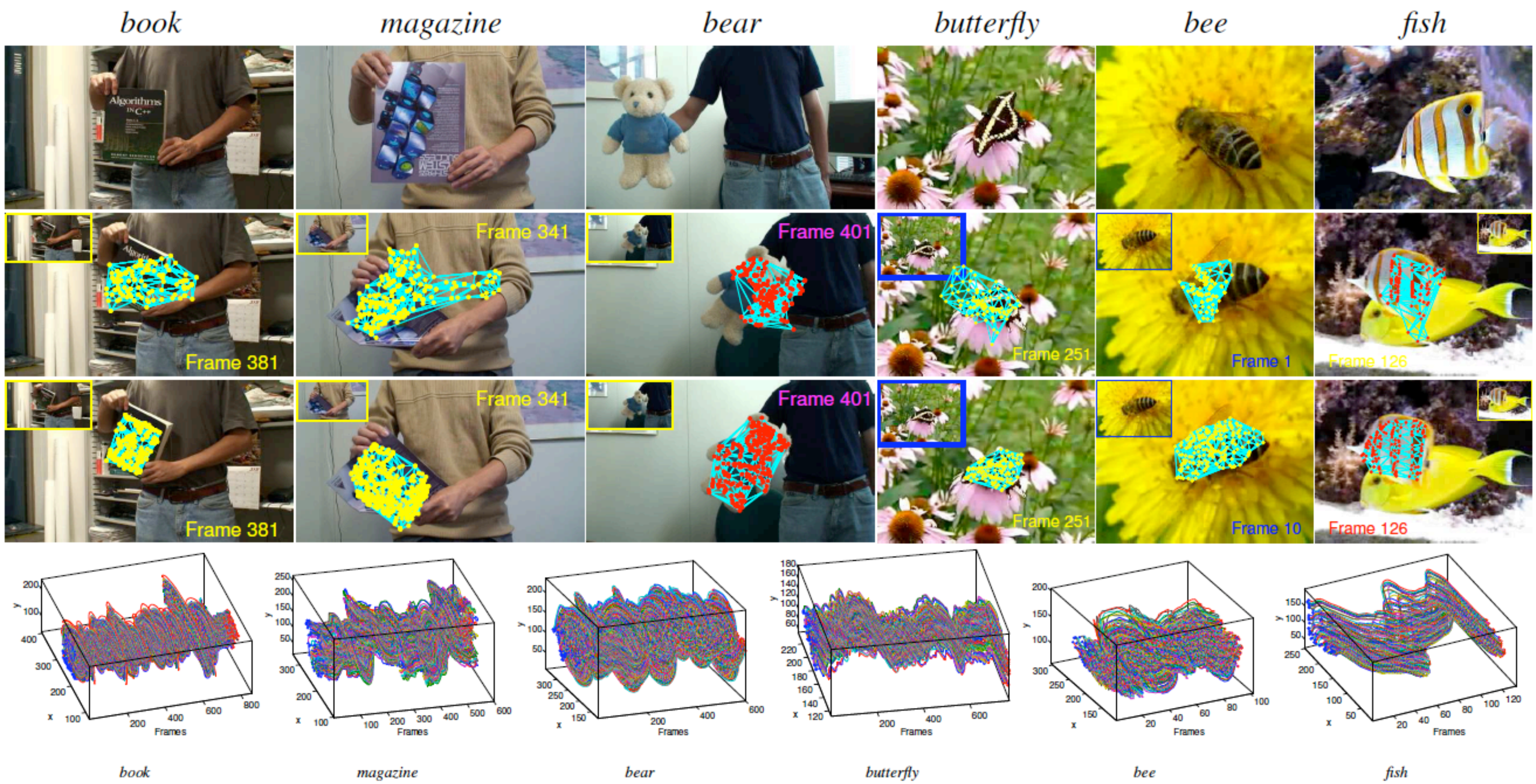




Statistics

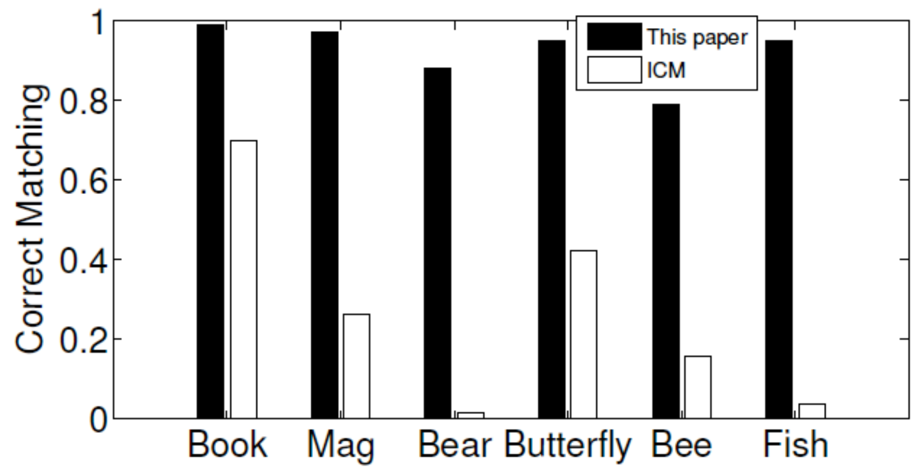
	book	magazine	bear	butterfly	bee	fish
#frames	856	601	601	771	101	131
#model	151	409	235	124	206	130
#target	2143	1724	1683	1405	1029	7316
time	1.6s	11s	2.2s	1s	2s	0.9s
accuracy	99%	97%	88%	95%	79%	95%

Results Comparison

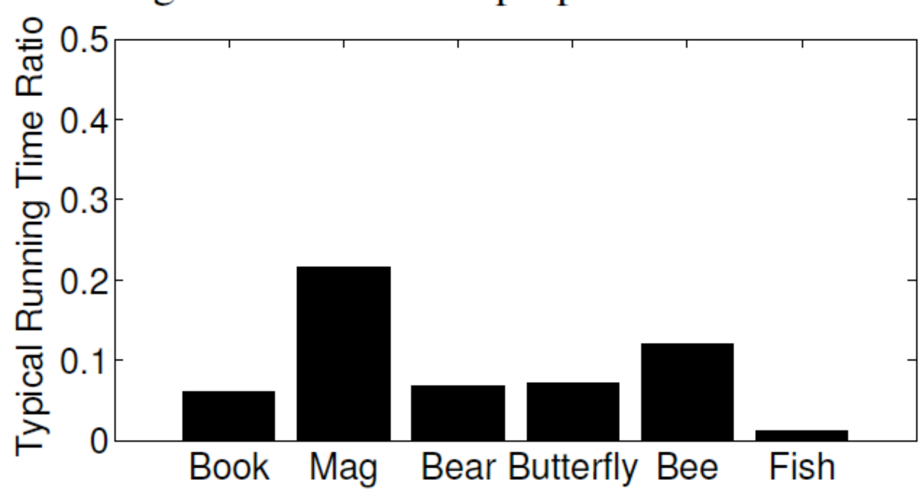


Accuracy and Efficiency

a) Detection rate in a video



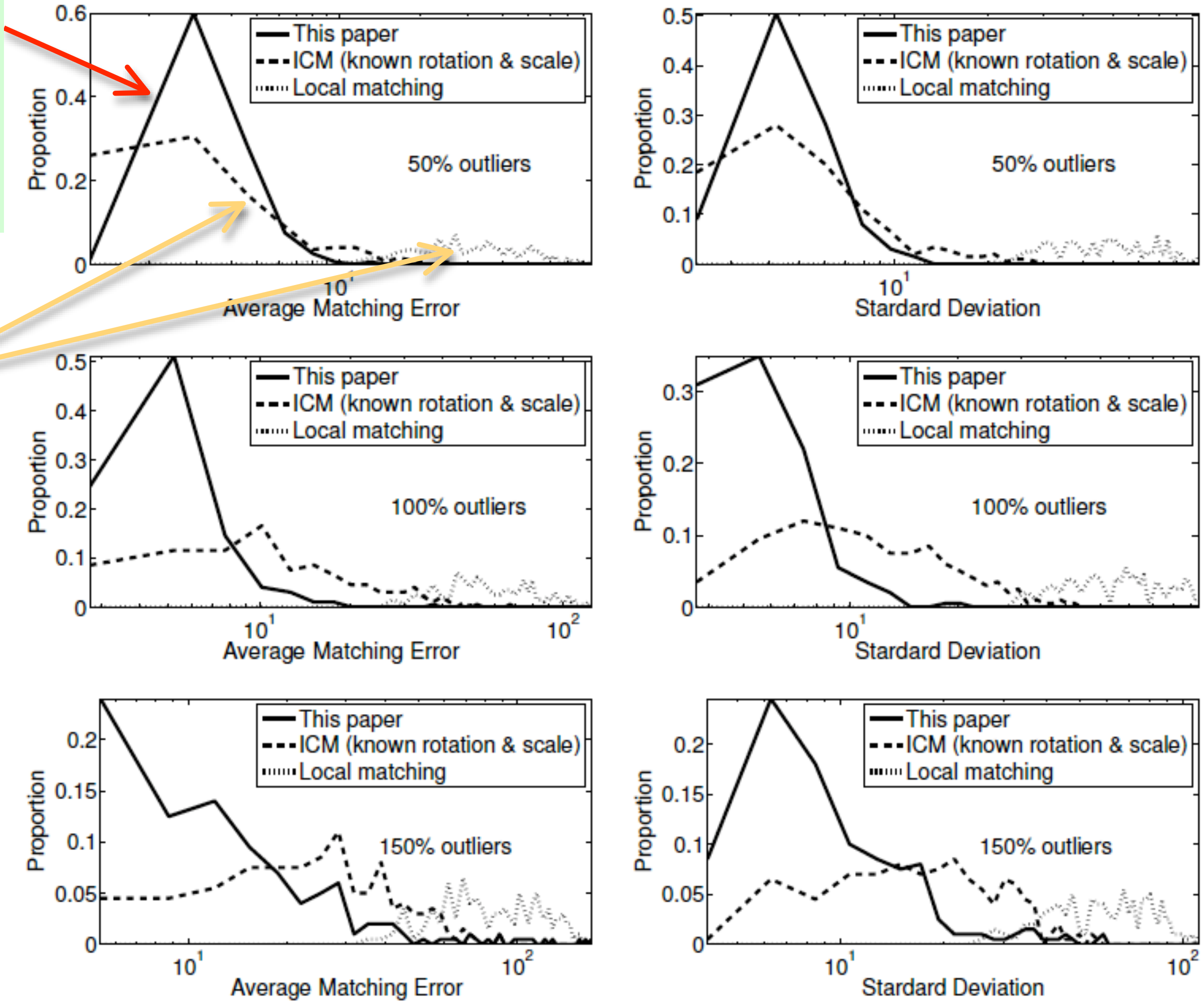
b) Running time ratio of the proposed method to ICM



Experiments on Ground Truth Data

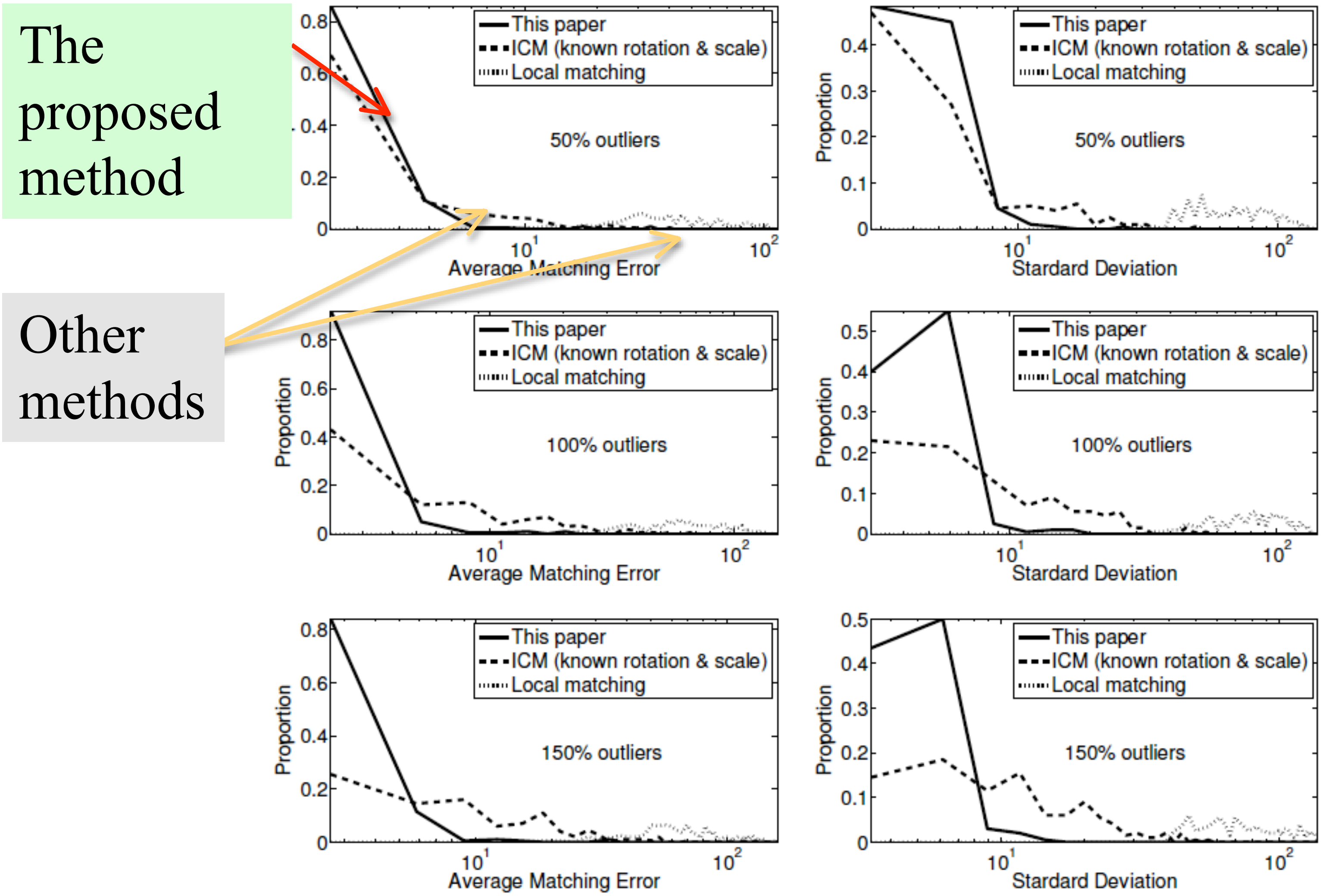
The proposed method

Other methods



Error distributions for fish dataset

Experiments on Ground Truth Data



Error distributions for random dot dataset

Summary

- A linear method to solve scale and rotation invariant matching problems accurately and efficiently
  - The proposed method is flexible and can be used to match images using different features
  - It is useful for many applications including object detection, tracking and activity recognition
- Future Work
  - Multiple scale solution
  - More complex transformations
  - Articulated object matching